基于变分自编码器的生成式文本摘要研究 *

黄佳佳[†]，李鹏伟

(南京审计大学 信息工程学院, 南京 211815)

摘　要：从单文档中生成简短精炼的摘要文本可有效缓解信息爆炸给人们带来的阅读压力。近年来，序列到序列(sequence-to-sequence，Seq2Seq)模型在各文本生成任务中广泛应用，其中结合注意力机制的 Seq2Seq 模型已成为生成式文本摘要的基本框架。然而，与机器翻译等任务不同，摘要文本还包含特定的写作风格特征。为生成能体现这种特征的摘要，在基于注意力和覆盖率机制的 Seq2Seq 模型基础上，在解码阶段利用变分自编码器(variational auto-encoder，VAE)刻画摘要风格特征并用于指导摘要文本生成；最后，利用指针生成网络来缓解模型中可能出现的未登录词问题。基于新浪微博 LCSTS 数据集的实验结果表明，提出的方法能有效刻画摘要风格特征、缓解未登录词及重复生成问题，使得生成的摘要准确性高于基准模型。

关键词：文本摘要；变分自编码器；Seq2Seq 模型；覆盖率机制；指针生成网络

中图分类号：TP399　　doi: 10.19734/j.issn.1001-3695.2020.03.0051

Vaebased abstractive text summarization

Huang Jiajia[†], Li Pengwei

(*School of Information Engineering*, *Nanjing Audit University*, *Nanjing* 211815, *China*)

Abstract: Generate a BRIEF and short summary for a document can effectively alleviate the reading pressure brought by information explosion. in recent years, sequence-to-sequence (seq2seq) model has been widely applied in various text generation tasks. especially, seq2seq attentional model has become a basic framework for abstractive text summarization task. however, different to other text generation task, such AS machine translation, summary text usually implies latent feature of writing style. to make the generated summary reflect the feature, this paper propose a novel framework, which employs seq2seq attentional model equipped with coverage mechanism AS basic model, and uses variational auto-encoder (vae) to depict the latent feature in decoding process. AT last, the summary tokens are generated successively via a pointer-generator network for relieving the out-of-vocabulary (OOV) problem. experimental results on the lcsts dataset demonstrate that the proposed framework is able to capture the latent feature of summary text and relief the OOV and tokens repetition problem, thus generating more accurate and readable summary.

Key words: text summarization; variational auto-encoder; Seq2Seq model; coverage mechanism; pointer-generator network

0　引言

随着互联网信息的爆发式增长，人们每天接触到海量信息，包括新闻、用户自媒体内容、聊天文本等。单文档自动文本摘要旨在从较长文本(如新闻、微博等)中提取出重要信息并以简短的句子表达之，以缓解信息过载所造成的阅读压力。然而，如何从文本中识别出重要信息，以及如何组织文字以表达这些重要信息是自动摘要所关注的两个基本问题。目前，文本自动摘要主要有抽取式摘要和生成式摘要两种方式。其中，抽取式摘要旨在从原始文本中找出若干个重要句子并作为摘要输出，一般用在长文本中；而生成式摘要旨在基于原始文本的语义信息自动生成语义连贯的简短句子。相比于抽取式摘要，生成式摘要更加符合人类的语言认知习惯，但如何使得生成的摘要文本尽可能涵盖原文核心信息且流畅自然是目前面临的主要挑战。

当前，生成式摘要主要采用包含注意力机制[1] (attention mechanism)的序列到序列(sequence-to-sequence，Seq2Seq)模型[2, 3]实现。而添加覆盖率机制[4](coverage mechanism)使得模型能够记录历史注意力分配，进而缓解摘要文本部分片段的重复生成问题。此外，由于模型在使用时存在未登录词(out-of-vocabulary，OOV)情况，使其应用效果不够理想。为此，相关学者在注意力模型基础上提出了指针生成网络[4, 5] (pointer-generator network)以缓解 OOV 问题，从而产生更加自然流畅的摘要句子。

结合注意力机制的 Seq2Seq 模型已成为单文档生成式摘要的基准方法。然而，相比于机器翻译等灵活多变的序列转换任务，自动摘要任务一般针对事件文本，因而往往具有潜在的写作风格特征，如“A 在 X 时间做了 B 事情”、“C 地点发生了 B 事情”等(如例 1 所示)。

例 1　具有结构特征的摘要文本示例

原始文本：今天下午，北京市…不再新建经济适用住房。
摘要：4000 元经适房今年退出北京历史
原始文本：雅虎发布 2014 年…至 51.45 美元。
摘要：雅虎宣布剥离阿里巴巴股份
原始文本：有着“全国最大包工头”称呼 … 严介说。
摘要：最大包工头严介和讨薪：状告地方政府拖欠工程款
原始文本：截至 10 月 28 日 … 楼市去库存速度明显提升。
摘要：91 家房企前三季度存货近万亿元

为此，Li 等人[6]在注意力模型[2]基础上提出一种考虑潜在结构信息的文本摘要生成模型 DRGD (deep recurrent

---

收稿日期：2020-03-20；**修回日期：**2020-04-27　**基金项目：**国家自然科学基金资助项目(61802194,61902190)；江苏省高等学校自然科学研究项目(17KJB520015,19KJB520040)

作者简介：黄佳佳(1989-)，女(通信作者)，安徽六安人，讲师，博士，主要研究方向为自然语言处理、审计数据分析(huangjj@nau.edu.cn)；李鹏伟(1987-)，男，讲师，博士，主要研究方向为软件安全与数据分析.

generative decoder)。该模型在解码阶段使用变分自编码器[7](variational auto-encoder, VAE)刻画文本摘要的结构特征并辅助生成摘要文本。然而该模型并未考虑自动摘要任务中的 OOV 和部分信息重复生成问题，产生的摘要质量有待进一步提高。因此，本文提出一种融合覆盖率机制、指针生成网络与摘要潜在风格特征的摘要生成模型 VAESum (VAE-based summarization generator)。具体而言，本文主要贡献如下：

a) 本文在覆盖率模型[4]基础上引入 VAE 网络刻画摘要文本的潜在风格特征，使得模型在解码阶段不仅考虑编码的隐藏层特征和拷贝原始文本的概率，还需要考虑摘要所包含的潜在风格信息，从而生成更高质量的摘要文本。

b) 本文基于新浪微博数据集 LCSTS[8]，对比分析了潜在风格特征与指针生成网络和覆盖率机制结合对摘要结果的影响。实验结果表明，当三者结合时产生的摘要质量最佳，并优于未考虑风格特征的摘要模型。

1　相关工作

随着深度学习在自然语言处理领域的推广与渗透，基于循环神经网络(recurrent neural network，RNN)的 Seq2Seq 模型[2]已成为生成式摘要任务的基本框架，并广泛应用于句子级别的摘要生成任务中，如新闻标题生成，摘要句子生成等。Seq2Seq 模型以原始文本的字符(或词)为输入特征，通过编码网络(encoder) 转换为隐藏层向量并传递到解码网络(decoder)，最终解码为摘要句子。在 Seq2Seq 模型中，为使解码网络更专注于输入文本中的重要特征，往往采用注意力机制[1, 3]来计算原始文本中每个特征对当前解码特征的贡献度，以期生成更加恰当的解码特征。在注意力模型基础上，Kikuchi 等人[9]考虑摘要长度，从而生成指定长度的摘要。

基于注意力机制的 Seq2Seq 模型存在重复生成问题，即生成的摘要片段可能存在部分字符重复出现，这是因为注意力机制往往不曾关注其历史分配情况。为此，相关研究提出解码器内部注意力[10](intra-decoder attention)、覆盖率机制[4, 11]或时序注意力机制[5](temporal attention)等以缓解重复生成问题。这些方法均考虑历史时刻的注意力分配使得当前注意力更加关注之前未被关注到的编码信息。

与机器翻译任务类似，自动摘要任务中常出现专有名词，这些名词在模型训练时不曾出现，但对当前文本的摘要生成却必不可少，进而导致未登录词(OOV)的现象出现。针对这一问题，Gu 等人[12]在注意力模型基础上提出拷贝网络(CopyNet)，即解码网络的输出层是由其激活层计算的摘要特征生成概率和该特征是否拷贝自原始文本的概率共同构成。拷贝网络显著提高了摘要质量，有效缓解了 OOV 问题。在此基础上提出的指针生成网络[4, 5]进一步考虑了词汇由词表产生的概率，而连续拷贝机制[13]可从输入文本中直接复制一个人名或机构名等文本子序列。在拷贝网络和指针生成网络模型中，在解码阶段的各个时刻，摘要字符不再完全由解码网络独立生成，而是考虑了将原始文本某个特征直接拷贝到摘要中的概率。

由于摘要自动生成任务的输入文本往往包含较多字符(一般大于 100 个)，而产生的摘要文本相对较短且包含较为固定的写作风格。为能够刻画摘要中所包含的潜在结构信息(即摘要风格)，Li 等人[6]在基于注意力的 Seq2Seq 模型中结合 VAE 网络将结构信息刻画出来并融入到摘要生成过程中。此外，针对新闻等长度较长的文本，相关研究将输入文本的主题词[14]、主题向量[15]等融入到编码器中以更加准确地概括输入文本的主要信息。

本文针对文本摘要自动生成任务面临的 OOV 和重复生成问题，以包含注意力、覆盖率机制以及指针生成网络的 Seq2Seq 模型为基准，在解码器中使用 VAE 网络刻画摘要文本的写作风格特征并融入到摘要循环生成过程中。本文提出的模型在尽可能缓解 OOV 和重复生成的同时，产生尽可能符合标准结构的摘要文本。

2　基于 VAE 的序列到序列模型

基于注意力、覆盖率机制和指针生成网络的 Seq2Seq 模型能较好解决摘要文本中出现的 OOV 问题以及重复生成问题。但从摘要角度，摘要文本由于其简练性要求，往往具有一定的写作风格特征。为此，本文在 Seq2Seq 模型基础上提出使用变分自编码器(VAE)刻画摘要文本的风格特征，并提出一种新的摘要生成模型 VAESum。

2.1　传统 VAE 结构

Kingma 等人[7]提出的变分自编码器是一种可利用潜在向量刻画数据潜在特征的编码-解码网络。在 VAE 模型的编码阶段，利用神经网络将输入数据 $\boldsymbol{d}$ 转换为潜在特征空间中的分布 $p(\boldsymbol{z}|\boldsymbol{d}) \sim \mathrm{N}(\boldsymbol{z};\boldsymbol{\mu}_d,\boldsymbol{\sigma}_d^2\boldsymbol{I})$，其中 $\boldsymbol{\mu}_d = \mathrm{f}_1(\boldsymbol{d})$、$\log\boldsymbol{\sigma}_d^2 = \mathrm{f}_2(\boldsymbol{d})$ 均由神经网络产生；在解码阶段，从 $\boldsymbol{z}$ 中采样并利用神经网络重构出原始数据 $\hat{\boldsymbol{d}}$。

经典的 VAE 模型并未应用于序列生成任务。随后 Chuang 等人[16]和 Li 等人[6]开始将 VAE 模型引入到 RNN 中，构建具有潜在特征循环生成能力的 Seq2Seq 模型，并应用于语音生成、手写字识别和文本摘要等任务。本文借鉴 DGRD[6]中的循环 VAE 模型作为 VAESum 模型解码网络的一部分。但与 DGRD 不同，本文不仅使用包含注意力机制的 Seq2Seq 作为基准生成模型，而且在此基础上进一步结合指针生成网络和覆盖率机制以缓解模型的 OOV 和重复生成问题。

2.2　提出的模型

本文提出的 VAESum 模型框架如图 1 所示。在编码器中，每个文本 $X = \{\boldsymbol{x}_1,\boldsymbol{x}_2,\ldots,\boldsymbol{x}_n\}$ 表达成词向量形式并输入到包含 GRU 单元(gated recurrent unit)的双向循环神经网络中生成隐藏层 $\vec{\boldsymbol{s}}_i = \mathrm{GRU}(\boldsymbol{x}_i,\vec{\boldsymbol{s}}_{i-1})$ 和 $\overleftarrow{\boldsymbol{s}}_i = \mathrm{GRU}(\boldsymbol{x}_i,\overleftarrow{\boldsymbol{s}}_{i-1})$，并拼接成最终的隐藏层向量 $\boldsymbol{s}_i = \vec{\boldsymbol{s}}_i \,\|\, \overleftarrow{\boldsymbol{s}}_i$。

解码器网络包含三个叠加子网络：a) 基于注意力和覆盖率机制的网络(ACM network, attention and coverage mechanism based network)；b) 基于 VAE 结构的网络(VAE network)；c) 基于指针生成网络的输出层(pointer-generator network)。三个子网络自底向上，依次以下层输出作为上层输入，最终生成摘要。

2.2.1 ACM 网络

ACM Network 使用两层神经网络刻画在注意力和覆盖率机制下如何生成解码器的隐藏层单元。具体地，首先利用编码器的输出隐藏层来初始化解码器网络的隐藏层 $\boldsymbol{h}_0 = \frac{1}{n}\sum_{i=1}^{n}\boldsymbol{s}_i$。随后，利用 t-1 时刻生成的字符向量 $\boldsymbol{y}_{t-1}$ 以及隐藏层单元 $\boldsymbol{h}_{t-1}^1$ 生成 t 时刻的第一隐藏层单元：

$$\boldsymbol{h}_t^1 = \mathrm{GRU}(\boldsymbol{y}_{t-1},\boldsymbol{h}_{t-1}^1) \tag{1}$$

值得注意的是，在模型训练时，使用 t-1 时刻的真实字符作为 t 时刻的输入；而当模型用于预测时，使用 t-1 时刻预测出的字符作为 t 时刻的输入。

编码器将输入文本转换成一个隐藏层向量传递给解码器，但解码器的第一隐藏层 $\boldsymbol{h}^1$ 仅使用编码器的信息为输入，而未考虑各时刻的解码单元是否需要关注输入文本的不同部分。实际上，在摘要生成任务中，若希望能够生成既可表达输入文本全局信息且重复字符较少的摘要文本，应要求解码器每个时刻只关注输入文本的部分信息，且在下一时刻减少对该部分的关注。Seq2Seq 模型一般使用注意力机制[3]使得解码器每个时刻只选择性地关注输入文本的部分信息；使用覆盖率机制[4]、时序注意力[5]或解码器内部注意力[10]刻画注意力向量的历史分布情况，以缓解重复生成和只关注部分文本的问题。

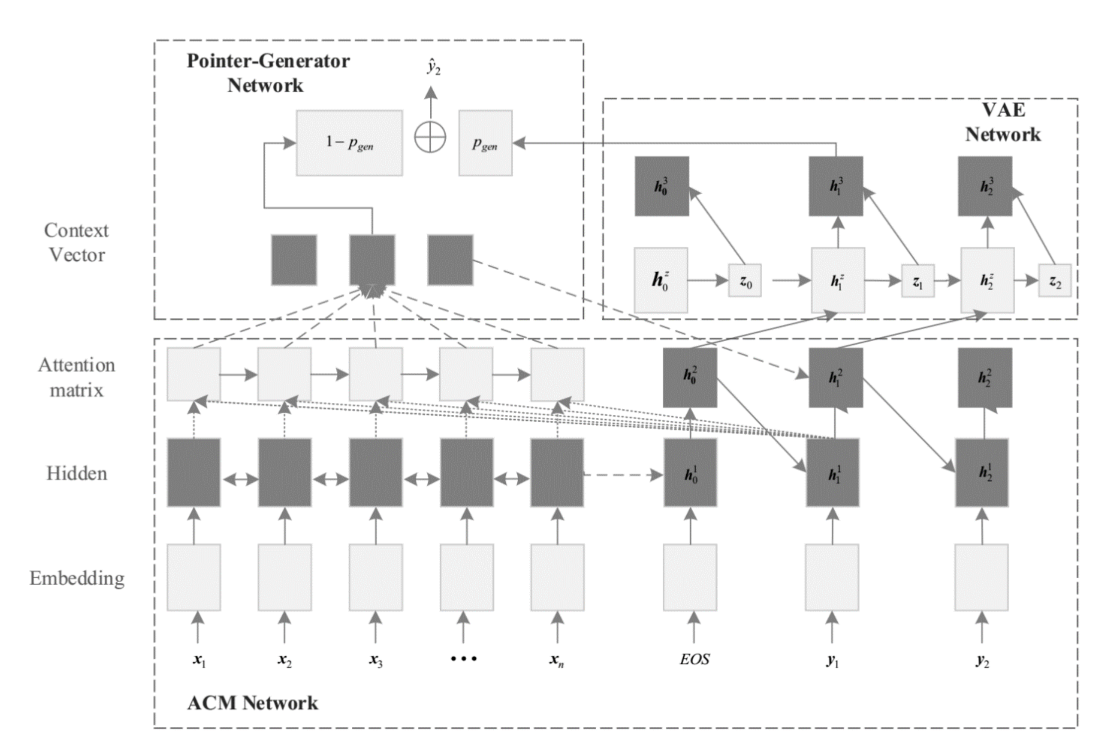


图 1 VAESum 网络结构图

Fig. 1 The vaesum network

本文使用注意力机制和覆盖率机制共同作用在解码器的 ACM network 隐藏层和编码器的隐藏层单元上。对于 ACM network 的两个隐藏层 h^1 和 h^2 来说，依次使用 t-1 时刻的第二隐藏层 h_{t-1}^2 来计算 t 时刻的第一隐藏层，即 $h_t^1=\mathrm{GRU}(y_{t-1},h_{t-1}^2)$，并以此计算 h_t^2。

具体来说，首先置 0 时刻的注意力向量为零向量，即 $\alpha_0=0$。那么在计算 t 时刻注意力时，首先考虑历史时刻的注意力分配以尽可能降低字符的重复生成。为此，引入覆盖率向量 c_t 计算到 t 时刻为止，历史注意力的累积覆盖情况：

$$c_t=\sum_{j=0}^{t-1}\alpha_j \tag{2}$$

根据 s_i、h_t^1 和覆盖率向量 c_t，解码器在 t 时刻对第 i 个输入文本的注意力为

$$e_t^i=v^T\tanh(W_s s_i+W_h^1 h_t^1+W_c c_t+b_{att}) \tag{3}$$

$$\alpha_t^i=\frac{\exp(e_t^i)}{\sum_{i'=1}^{n}\exp(e_t^{i'})} \tag{4}$$

结合该注意力以及编码器的隐藏层 s_i 可获得在考虑注意力和覆盖率情况下解码器在 t 时刻应该关注的上下文向量：

$$C_t=\sum_{i=1}^{n}\alpha_t^i s_i \tag{5}$$

这样，综合考虑 t-1 时刻的输入字符向量 y_{t-1} 以及 t 时刻第一隐藏层 h_t^1、上下文向量 C_t，可获得解码器在 t 时刻的第二隐藏层：

$$h_t^2=\mathrm{GRU}(y_{t-1},h_t^1,C_t) \tag{6}$$

2.2.2 VAE 网络

一般的 Seq2Seq 模型较少考虑摘要文本的潜在写作风格特征。而文献[6]的实验结果表明，结合 VAE 网络的 Seq2Seq 模型将摘要的潜在风格特征考虑进来可有效提升摘要质量。鉴于此，本文在包含注意力和覆盖率机制的解码器模块上进一步融入 VAE 网络以刻画摘要风格特征。

VAE 网络以 ACM 网络的第二隐藏层 h^2 为输入，由于各时刻的隐藏层之间包含上下文语义联系，因而不能使用原始 VAE 网络独立地为每时刻 h_t^2 生成下一层节点。与 DGRD[7]中的循环 VAE 生成过程类似，给定 ACM 网络 t 时刻输出的隐藏层 h^2 和 t-1 时刻的输入字符向量 y_{t-1}，首先生成当前网络的潜在隐藏层：

$$h_t^z=\mathrm{f}(W_y^z y_{t-1}+W_{hh}^z h_{t-1}^2+W_{zh}^z z_{t-1}+b_z) \tag{7}$$

其中，f 为 sigmoid 函数，z_t 为 t 时刻的潜在特征，该特征刻画了摘要文本的风格信息。

VAE 模型假设 t 时刻潜在特征满足高斯先验 $z_t\sim\mathrm{N}(\boldsymbol{0},\boldsymbol{I})$，且每个隐藏层在 z 上的后验分布也满足高斯分布 $p(h_t^z|z_t)\sim\mathrm{N}(z_t;\mu_t,\sigma_t^2\boldsymbol{I})$，其中后验分布参数 μ_t 和 $\log\sigma_t^2$ 由神经网络生成：

$$\mu_t=W_\mu h_t^z+b_\mu \tag{8}$$

$$\log\sigma_t^2=W_\sigma h_t^z+b_\sigma \tag{9}$$

利用重参数化技巧(reparameterization trick)即可刻画出 t 时刻的潜在特征 z_t：

$$z_t=\mu_t+\sigma_t\otimes\varepsilon,\quad \varepsilon\sim\mathrm{N}(\boldsymbol{0},\boldsymbol{I}) \tag{10}$$

最后，根据潜在特征 z_t 获得 VAE 网络的输出隐藏层：

$$h_t^3=\tanh(W_{zh}z_t+W_{hh}^3 h_t^2+b_h^3) \tag{11}$$

该隐藏层综合考虑了摘要文本的潜在风格特征以及输入文本的注意力信息。

2.2.3 pointer-generator 网络

虽然对 VAE 网络输出的隐藏层 h_t^3 使用 softmax 函数直接映射为字符输出层单元即可生成摘要，但这种方式未考虑模型预测时出现的未登录词问题。即若输入文本 X 中的某个字符或词汇不在模型的训练词表中，那么上述方式将无法生成相应字符或词汇。因此，本文采用指针生成网络来决定 t 时刻预测出的字符是直接从词表中产生还是复制自输入文本的某个字符。

为此，首先将影响 t 时刻字符生成的相关向量拼接起来，包括 t-1 时刻生成的字符向量 y_{t-1}、VAE 网络的输出隐藏层 h_t^3、当前时刻的上下文向量 C_t：$h_t=[y_{t-1},h_t^3,C_t]$。然后计算字符从词表中产生的概率 p_{gen} 以及每个字符 w 被选中的概率 $p_{voc}(w)$：

$$p_{gen}=\mathrm{sigmoid}(W_t h_t+b) \tag{12}$$

$$p_{voc}(w)=\mathrm{softmax}(W_t' h_t+b') \tag{13}$$

这样，字符 w 从词表中产生的最终概率为 $p_{voc}(w)p_{gen}$。若字符 w 不在词表中，那么 $p_{voc}(w)$ 为 0，这时利用输入文本中每个字符 w_i 在 t 时刻的注意力 α_t^i 来从输入文本中复制最合适的字符。模型预测出每个字符的概率为

$$p(w)=p_{voc}(w)p_{gen}+(1-p_{gen})\sum\nolimits_{w_i=w}\alpha_t^i \tag{14}$$

根据式(14)可逐字符生成摘要文本，该过程直到产生停止字符<EOS>或摘要文本达到指定最大长度时停止。此外，为提高模型预测效率，本文使用束搜索[17](beam search)方式来生成最佳摘要。

2.3　模型训练

本文提出的VAESum文本摘要模型是在基于注意力和覆盖率机制的 Seq2Seq 模型基础上利用 VAE 网络刻画摘要文本的潜在结构并融入到摘要生成任务中。因此，为训练模型参数，首先利用交叉熵[18](cross-entropy)来最小化长度为 T 的真实摘要 $\hat{Y}^j=\{y_1^j,y_2^j,\ldots,y_T^j\}$ 在每一时刻的对数似然概率：

$$\ell_E=-\log\frac{1}{T}\sum_{t=1}^{T}p(y_t^j\mid y_{<t}^j,X) \tag{15}$$

其次，本文使用覆盖率机制来度量历史注意力分布(即覆盖率向量)并以此优化当前时刻的注意力分配。根据文献[4]的实验结果，将覆盖率损失函数 $\ell_c=\sum_i\min(\alpha_t^i,c_t^i)$ 添加到优化函数中可有效降低注意力对某个文本片段的重复注意问题。因此，本文在模型训练时也将该损失函数作为优化函数一部分。

最后，VAE 网络是一个无监督的生成-推断网络。其模型参数的训练方式如下：在最大化每个摘要每一时刻生成概率 $p(y_{<t}^j)$ 的同时尽可能使得从文本中训练出的后验概率 $q(z_t^j\mid y_{<t}^j,z_{<t}^j)$ 逼近其理论变分概率 $p(z_t^j\mid y_{<t}^j,z_{<t}^j)$。这样，其对应优化函数的最终表达式如下[6, 16]：

$$\begin{aligned}\ell_V=E_{q(z_t^j\mid y_{<t}^j,z_{<t}^j)}\Big\{&\sum\nolimits_{t=1}^{T}\log p(y_t^j\mid y_{<t}^j,z_t^j,X)\\&-D_{KL}[q(z_t^j\mid y_{<t}^j,z_{<t}^j)\parallel p(z_t^j)]\Big\}\end{aligned} \tag{16}$$

由于 ℓ_E 与 ℓ_V 均包含摘要文本生成概率的对数似然优化，可将两部分合并。这样，对于具有 N 个文本摘要对的训练数据 $\{X,Y\}_N$，本文提出的 VAESum 模型的整体优化函数如下：

$$\begin{aligned}\ell=\frac{1}{N}\sum_{n=1}^{N}\sum_{t=1}^{T}\{&-\log p(y_t^j\mid y_{<t}^j,X)+\lambda\ell_c\\&+D_{KL}[q(z_t^j\mid y_{<t}^j,z_{<t}^j)\parallel p(z_t^j)]\}\end{aligned} \tag{17}$$

其中，λ 为覆盖率损失的调节参数。

3　实验

3.1　数据集

本文实验采用的是 Hu 等人[8]提供的新浪微博数据集 LSCST(large-scale chinese short text summarization dataset)。该数据集以微博短文及其摘要作为文本-摘要对。整个数据集分为训练、验证和测试三部分。为更有效地评估摘要模型，对验证数据和测试数据进行人工打分(1-5 分)，并保留分数不低于 3 分的数据。最终，各部分分别包含约 240 万、8700 和 725 个文本—摘要对。

3.2　实验设置与对比模型

预处理时以文本字符流为输入，这是因为若干研究结013表明，基于中文字符的摘要模型效果更佳[8, 15]。在模型参数设置方面，最大文本长度和最大摘要长度分别为 120 和 25，词典大小为 4000 个字符；词向量为 350 维，潜在特征和隐藏层的维度均为 500 维；批大小为 256，束搜索范围为 10，覆盖率损失参数 λ=1.0。本文采用 AdaDelta[19]方法进行梯度下降训练模型参数，其中学习率为 0.5。本文在 Pytorch 框架上实现模型代码，以 NVIDIA Tesla 加速训练，模型共训练 33 个 epochs，耗时约 8 天。

本文将与以下使用 LSCST 数据集的基准模型对比，并从相关文献中直接抽取实验结果。

a) RNN 和 RNN-context[8]：即提出 LSCST 短文本摘要数据集的两个基于 RNN 的文本摘要模型。其中 RNN-context 模型中使用了注意力机制。

b) CopyNet+W[12]：一种使用拷贝网络的 Seq2Seq 模型，并使用文本的词序列为输入。

c) DRGD[6]：一种基于注意力机制的 Seq2Seq 模型，并在解码阶段添加 VAE 网络刻画摘要文本的潜在特征。

d) Cover-5[11]：一种基于注意力和覆盖率机制的 Seq2Seq 文本摘要模型。

e) PGC[4]：一种基于注意力、覆盖率和指针生成网络的 Seq2Seq 摘要模型。该模型也是本文的基准模型。由于该模型未在 LSCST 数据集上测试，本文使用文献提供的代码(www.github.com/abisee/pointer-generator)及文献中设置的实验参数在 LSCST 上训练模型。

3.3　实验结果

3.3.1 ROUGE evaluation

本文首先使用 Lin 等人[21]提出的 ROUGE(recall-oriented understudy for gisting evaluation)指标对比评估各模型。该指标以生成的摘要在标准摘要中 n-元公共子序列个数来评价摘要模型的优劣，其中 R-1 和 R-2 分别指 1-元和 2-元子序列，R-L 指最长公共子序列。

表 1 对比分析了各种文本摘要模型在 LCSTS 数据集上的实验结果。从表 1 中可以看出，相比于其他基准模型和 VAESum 模型的弱化形式，本文提出的 VAESum 模型在 ROUGE 三个指标上均有一定程度的提升。

表 1　ROUGE 评估结果对比

Tab. 1　ROUGE Results on LSCST

方法	R-1	R-2	R-L
RNN	21.5	8.9	18.6
RNN-context	29.9	17.4	27.2
CopyNet+W	35.0	22.3	32.0
Cover-5	33.58	21.00	31.21
PGC	34.46	21.09	32.02
DRGD	36.99	24.15	34.23
VAESum-Cov[1]	37.24	24.18	34.26
VAESum-Copy[2]	37.45	24.25	34.44
VAESum	37.74	24.87	34.80

注：a)VAESum-Cov 是指在 VAESum 模型的 ACM 子网络中未使用覆盖率机制来计算注意力分布；

b)VAESum-Copy 是指在 VAESum 模型中未使用指针生成网络，而是根据式(13)直接生成摘要字符。

此外，在解码网络中添加 VAE 模块能够有效捕获摘要文本的写作风格信息。因此，相比于仅使用注意力机制和拷贝机制(或覆盖率机制)的模型(如 RNN-context、PGC)，添加 VAE 结构的摘要模型(如 DRGD、VAESum)的性能比其基准模型均有较为显著的提升。

3.3.2 案例展示

为进一步直观评估 VAESum 模型的摘要生成能力，表 2 展示了若干个摘要结果样例。从表 2 中可以看出，相比于 PGC 模型，在解码网络中添加 VAE 模块能够有效捕获摘要文本的写作风格信息，特别是基于 PGC 网络并添加 VAE 模块的 VAESum 模型所生成的摘要句子句法结构和语义信息最为完整，且表达了微博文本的主要含义。

此外，相比于指针生成网络，考虑历史注意力的覆盖率机制对摘要生成质量的影响更大。当 VAESum 模型未使用覆盖率机制(即 VAESum-Cov)时，产生的摘要往往只能表达微博文本的部分信息，较难涵盖全部信息。而当未使用指针生成网络(即 VAESum-Copy)时，产生的摘要基本表达了微博文本的全部信息，但常遗漏专有名词的部分片段，如样例 1 中的“中联航空”只生成“中联航”；样例 2 中“广州军区”只生成“广州”等。

表 2 摘要结果样例

Tab. 2 Examples of the generated summaries

模型	文本数据
微博文本(1)	昨晚，中联航空成都飞北京一架航班被发现有多人吸烟。后因天气原因，飞机备降太原机场。几名乘客在舱门边吸烟被发现。有乘客要求重新安检，机长决定继续飞行，引起机组人员与未吸烟乘客冲突。目前中联航空正联系机组进行核实。
参考摘要	成都飞北京航班多人吸烟机组人员与未吸烟乘客冲突
PGC	中联航空正联系机组进行核实
VAESum-Cov	中联航空成都飞北京航班备降太原机场
VAESum-Copy	中联航成都飞北京航班多人吸烟
VAESum	中联航空成都飞北京航班遭多人吸烟
微博文本(2)	昨日上午#时，广州军区空军某部在组织正常飞行训练时，一架歼#飞机在起飞上升过程中突发机械故障，飞机状态无法控制，坠落在广东省汕头市郊区，飞行员跳伞成功。地面#名受伤群众被第一时间送到医院治疗。
参考摘要	一架歼#飞机训练时在汕头坠毁
PGC	广州军区空军某部正常飞行训练时突发机械故
VAESum-Cov	广州军区一架歼飞机起飞突发机械故障
VAESum-Copy	广东一架歼飞机突发故障坠落广东汕头飞行员
VAESum	广州军区一架歼飞机起飞上升过程中突发故障
微博文本(3)	#月#日，全国性地方债审计全面开闸。审计的背后，是部分地方政府盲目举债的隐忧及无力还债的现实。媒体披露审计的#个目标省会城市中，债务压力排名前#为：南京、成都、广州、合肥、昆明、长沙、武汉、哈尔滨、西安和兰州。
参考摘要	媒体公布内地省会中债务压力排名最高城市名单
PGC	全国性地方债审计全面开闸部分地方政府盲目
VAESum-Cov	全国性地方债审计全面开闸部分地方盲目举债
VAESum-Copy	审计报告称省会城市中债务压力排名前
VAESum	省会城市中债务压力排名出炉
微博文本(4)	又一位券商固收高管被卷入“债市打黑”风暴。这是继国信、宏源等多家券商固定收益部负责人被调查之后的最新进展。随着“债券女王”孙明霞供出的一百多人名单的逐个排查，这一轮祸起发改委企业债链条的窝案或将结束。
参考摘要	发改委企业债链条打黑：上百人名单逐个排查
PGC	发改委企业债链条窝案或将结束
VAESum-Cov	发改委企业债链条窝案或结束
VAESum-Copy	券商固收高管被卷入债市打黑风暴
VAESum	券商固收高管被卷入债市打黑风暴或将结束

注：其中“#”指训练模型中的停用字符。

最后，纵观测试集中的全部摘要，本文提出的 VAESum 模型较难刻画带有标点符号的写作风格，如样例 4 所示的“XX：YY”风格。部分原因是预处理时将“：”“？”等符号置为停用字符，从而导致无法判断生成的摘要是否需要添加标点符号。例如，生成的摘要“券商固收高管被卷入债市打黑风暴或将结束”或可规范化为“券商固收高管被卷入‘债市打黑’风暴，(风暴)或将结束”或“券商固收高管被卷入，‘债市打黑’风暴或将结束”。

4 结束语

本文提出一种新的生成式单文本自动摘要模型 VAESum 以提升自动摘要性能。该方法在基于注意力机制的序列到序列模型基础上，首先采用覆盖率机制进一步优化摘要文本片段重复生成问题；其次，引入 VAE 网络刻画摘要文本的风格特征并融入到解码网络中；最后使用指针生成网络以缓解 OOV 问题。在 LCSTS 数据集上的实验结果表明，本文提出的 VAESum 模型比基准方法的摘要生成能力得到了提升。

参考文献：

[1] Chopra S, Auli M, Rush A M. Abstractive sentence summarization with attentive recurrent neural networks [C]// Proceedings of the 15th Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Pennsylvania: ACL Press, 2016: 93-98.

[2] Sutskever L, Vinyals O, Le. Q V. Sequence to sequence learning with neural networks [C]// Proceedings of the 27th Annual Conference on Neural Information Processing Systems, 2014: 3104-3112.

[3] Rush A M, Chopra S, Weston J, *et al.* A neural attention model for abstractive sentence summarization [C]// Proceedings of the 20th Conference on Empirical Methods in Natural Language Processing, Pennsylvania: ACL Press, 2015: 379-389

[4] See A, Liu P J, Manning C D, *et al.* Get to the point: summarization with pointer-generator networks [C]// Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics, Pennsylvania: ACL Press, 2017: 1073-1083.

[5] Nallapati R, Zhou Bowen, Santos C N, *et al.* Abstractive text summarization using sequence-to-sequence RNNs and beyond [C]// Proceedings of the 20th SIGNLL Conference on Computational Natural Language Learning, 2016: 280-290.

[6] Li Piji, Lam Wai, Bing Lidong, *et al.* Deep recurrent generative decoder for abstractive text summarization [C]// Proceedings of the 22th Conference on Empirical Methods in Natural Language Processing, Pennsylvania: ACL Press, 2017: 2091-2100.

[7] Kingma D P, Welling M. Auto-encoding variational Bayes [C]// Proceedings of the 2nd International Conference on Learning Representations, 2014.

[8] Hu Baotian, Chen Qingcai, Zhu Fangze, *et al.* LCSTS: a large scale Chinese short text summarization dataset [C]// Proceedings of the 20th Conference on Empirical Methods in Natural Language Processing, Pennsylvania: ACL Press, 2015: 1967-1972.

[9] Kikuchi Y, Neubig G, Sasano R, *et al.* Controlling output length in neural encoder-decoders [C]// Proceedings of the 21st Conference on Empirical

Methods in Natural Language Processing, Pennsylvania: ACL Press, 2016: 1328-1338.

[10] Paulus R, Xiong C, Socher R, *et al.* A deep reinforced model for abstractive summarization [C]// Proceedings of the 6th International Conference on Learning Representations, 2018.

[11] 巩轶凡，刘红岩，何军，等. 带有覆盖率机制的文本摘要模型研究 [J]. 计算机科学与探索，2019, 13 (2): 205-213. (Gong Yifan, Liu Hongyan. , He Jun. , et. al. Research on Text Summarization Model with Coverage Mechanism [J]. Journal of Frontiers of Computer Science and Technology. 2019, 13 (2): 205-213.)

[12] Gu Jiatao, Lu Zhengdong, Li Hang, *et al.* Incorporating copying mechanism in sequence-to-sequence learning [C]// Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics, Pennsylvania: ACL Press, 2016: 1631-1640.

[13] Zhou Qingyu, Yang Nan, Wei Furu, *et al.* Sequential copying networks [C]// Proceedings of the 32nd AAAI Conference on Artificial Intelligence, 2018: 4987-4995.

[14] 侯丽微，胡珀，曹雯琳. 主题关键词信息融合的中文生成式自动摘要研究 [J]. 自动化学报，2019, 45 (3): 530-539. (Hou Liwei, Hu Po, Chao Wenlin. Chinese abstractive summarization with topical keywords fusion [J]. Acta Automatica Sinica, 2019, 45 (3): 530-539.)

[15] Wang Li, Yao Junlin, Tao Yunzhe, *et al.* A reinforced topic-aware convolutional sequence-to-sequence model for abstractive text summarization [C]// Proceedings of the 27th International Joint Conference on Artificial Intelligence, 2018: 4453-4460.

[16] Chung J, Kastner K, Dinh L, *et al.* A recurrent latent variable model for sequential data [C]// Proceedings of the 28th Annual Conference on Neural Information Processing Systems, Pennsylvania: ACL Press, 2015: 2980-2988.

[17] Koehn P. Pharaoh. A beam search decoder for phrase-based statistical machine translation models [C]// Proceedings of the 6th Conference of the Association for Machine Translation in the Americas, Pennsylvania: ACL Press, 2004: 115-124.

[18] Ranzato M, Chopra S, Auli M, *et al.* Sequence level training with recurrent neural networks [C]// Proceedings of the 4th International Conference on Learning Representations, 2016.

[19] Schmidhuber J. Deep learning in neural networks [J]. Neural Networks, 2015: 85-117.

[20] Blei D M, Ng A Y, Jordan M I. Latent Dirichlet allocation [J]. Journal of Machine Learning Research, 2003, 3 (Jan): 993-1022.

[21] Lin C. ROUGE: a package for automatic evaluation of summaries [C]// Proceedings of the 42nd Annual Meeting of the Association for Computational Linguistics, Pennsylvania: ACL Press, 2004: 74-81.